Published in final edited form as:

Curr Opin Chem Biol. 2018 August ; 45: 10–17. doi:10.1016/j.cbpa.2018.01.016.

Harnessing natural DNA modifying activities for editing of the genome and epigenome

Jamie E. DeNizio¹, Emily K. Schutsky¹, Kiara N. Berrios¹, Monica Yun Liu¹, and Rahul M. Kohli^{2,3}

¹Graduate Group in Biochemistry and Molecular Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

²Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

³Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract

The introduction of site-specific DNA modifications to the genome or epigenome presents great opportunities for manipulating biological systems. Such changes are now possible through the combination of DNA-modifying enzymes with targeting modules, including dCas9, that can localize the enzymes to specific sites. In this review, we take a DNA modifying enzyme-centric view of recent advances. We highlight the variety of natural DNA-modifying enzymes – including DNA methyltransferases, oxygenases, deaminases, and glycosylases – that can be used for targeted editing and discuss how insights into the structure and function of these enzymes has further expanded editing potential by introducing enzyme variants with altered activities or by improving spatiotemporal control of modifications.

Graphical Abstract

Corresponding author: Kohli, Rahul M (rkohli@penmedicine.upenn.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

A primary goal in epigenetic research has been to understand how enzymatic modification of DNA can regulate different transcriptional programs in cells that share a common genome. Much of our understanding so far has come from analyzing the phenotypes that result after deleting the relevant DNA-modifying enzymes. However, the last several years have marked a revolutionary transition from global to site-specific analyses, whereby naturally-occurring enzymes can now be harnessed to target and synthetically manipulate genomes, potentially enabling more direct accounting of cause and effect at specific loci. Even as epigenomic editing seems poised to demystify the functions of DNA modifications, editing of the DNA sequence itself has also advanced swiftly, offering tools to understand or possibly reverse specific mutations. While these compelling topics have been addressed in several recent comprehensive reviews [1,2], in this short review we focus on the biochemistry and chemical biology of the DNA-modifying enzymes. After introducing nature's 'toolbox' for modifying DNA, we describe the creative approaches taken by biochemists and chemical biologists to target these enzymes to particular loci and to manipulate the enzymes to expand their functional potential or to achieve more precise spatiotemporal control. Based on recent developments in DNA-modifying enzyme biochemistry, we also highlight future directions which, while highly anticipated, have yet to be fully realized.

Nature's enzymatic toolbox for DNA modification

The majority of the mammalian 'toolbox' for DNA modifications is centered on cytosine bases (Figure 1) [3]. While DNA methylation in bacteria is more diverse, the most common chemical modification in mammalian genomes is 5-methylcytosine (5mC), predominantly found in CpG dinucleotide contexts. 5mC is generated by the maintenance DNA methyltransferase 1 (DNMT1) and *de novo* DNMT3 enzymes, which utilize the methyl donor *S*-adenosylmethionine [4]. While methylation signatures can be stable in cell lineages, dynamic changes are readily evident during differentiation, in early embryos, and in response to environmental cues. These dynamics are mediated in part by another 'tool' in the enzymatic modification toolbox: ten-eleven translocation (TET) family enzymes [5]. These Fe(II)/ α -ketoglutarate-dependent dioxygenases can oxidize 5mC to yield 5-hydroxymethylcytosine (5hmC), a modification whose levels can almost rival that of 5mC in some cell types such as neurons [5]. TET enzymes can further oxidize 5hmC to 5-formylcytosine (5fC) and then 5-carboxylcytosine (5caC), expanding the repertoire of natural modifications. While each of these oxidized 5mC bases (ox-mCs) potentially has independent functions, they all appear to play a role in the process of DNA demethylation. Ox-mC bases on one strand can decrease the maintenance activity of DNMT1 on the opposite strand and thus result in passive dilution of 5mC with replication [5]. Importantly, 5fC and 5caC can also mark sites for *active* demethylation by invoking another of nature's tools: DNA glycosylases [6]. DNA glycosylases recognize and excise specific nucleobases, resulting in the generation of abasic sites. In active demethylation, thymine DNA glycosylase (TDG) can excise 5fC and 5caC, and completion of base excision repair (BER) leads to replacement of the modified base by unmodified cytosine at these sites [5].

AID/APOBEC cytosine deaminases, which catalyze the hydrolytic deamination of cytosine to generate uracil [7], are an added tool for DNA modification. These deaminases can change the primary DNA sequence and may also crosstalk with epigenetic cytosine modifications. AID works in concert with BER machinery, including uracil DNA glycosylase (UDG), to mutate the immunoglobulin locus and promote antibody maturation, while APOBEC enzymes can have other targets including retroviral replication intermediates and RNA [7]. Deamination of modified cytosine bases by AID/APOBECs has also been implicated in DNA demethylation, as 5mC or ox-mC deamination could theoretically lead to genomic mismatches whose repair could result in installation of an unmodified cytosine [8]. While biochemical evidence suggests that this is unlikely a major contributor to demethylation [9,10], the proposed pathway offers an additional, unexploited avenue for potential manipulation of 5mC and ox-mCs in genomes.

The formula for an editing complex (TM + MM)

The power of site-specific genome and epigenome editing has come as a result of advances in the use of DNA targeting proteins, which can be partnered with natural and/or altered DNA-modifying enzymes. To provide a framework for understanding these methodological developments, we offer a generic definition of the editing complexes (Figure 2). The typical formula for an editing complex involves a DNA targeting module (TM) partnered with one or more DNA modifying modules (MM). As we describe in the subsequent sections, TMs can recognize specific sequences or even specific DNA modifications within those sequences, while MMs can be partnered with accessory modules (MM_X) or can be variants of the natural enzyme with altered biochemical properties (MM*) or with tunable activity (MM⁺).

Targeting modules (TM)

While potential TMs have been well reviewed recently [11,12], we briefly note the available constructs to frame our discussion of DNA-modifying enzymes. The original method for programmable targeting of DNA sequences employed zinc-finger proteins, referred to as ZFPs, which recognize a trinucleotide motif and can be concatenated to recognize longer stretches of DNA [13]. The discovery of transcription activator-like effector (TALE) proteins from plant pathogens later offered a means for targeting single bases, and, like with ZFPs, concatenation of TALEs results in selective recognition of longer DNA motifs [14]. Structure-function insights into the TALE residues that contact nucleobases have permitted the development of constructs that differentiate unmodified cytosine from 5mC and ox-mCs [15–18]. Such modified TMs (TM*) have not yet been employed with different MMs, but the combination would make for a logical and anticipated advance. While ZFPs and TALEs are still being employed, the CRISPR/Cas system offers a more adaptable solution to the challenges of designing TMs, in that different guide RNAs (gRNA) could be used to achieve targeting to multiple locations with a single protein scaffold, rather than having to manipulate individual proteins for each target. Although Cas9 was originally exploited for targeted DNA cleavage, using catalytically inactive Cas9 (dCas9) or nicking variants (nCAs9) allows for retention of localization capabilities and partnering with DNA-modifying enzymes as MMs. The discovery of new CRISPR/Cas systems will likely expand the TM

choices yet further [19]. Across the variety of combinations, TM and MMs can be directly fused, which requires optimization of module ordering and linker lengths, or can be brought together by exploiting protein-protein or protein-RNA interactions (in the case of Cas9) to promote co-localization.

Targeting native DNA-modifying enzymes (MM)

With a suite of TMs to achieve targeting, the natural enzymatic ‘tools’ have been employed as MMs for locus-specific editing over the last several years. With various isozymes or homologues available across different kingdoms of life, selection among natural enzymes can be dictated by the editing goals. Starting with DNA methylation, DNMT3A has been fused to ZFPs, TALEs, and dCas9 for targeted methylation [20–27] and is likely preferred over DNMT1 because DNMT3A is less sensitive to the modification status of the opposite strand. For targeted DNA demethylation, the catalytic domain of human TET1 has primarily been used thus far, coupled to ZFPs, TALEs, and more commonly now to dCas9 [27–33]. Intriguingly, plants have a direct demethylation pathway through the protein ROS1, a DNA glycosylase that directly excises 5mC. After fusion with a yeast Gal4-DNA binding domain as a TM, ROS1 has also been employed for direct demethylation and specific reactivation of silenced genes [34]. The ROS1 construct represents the concept of using a native MM out of its native context, and such strategies might be more generally employed; for example, the well-behaved TET homologue from *Naegleria gruberi* could be employed instead of mammalian TET enzymes [35], or the bacterial GpC methyltransferase M.CviPI could be used to assess methylation in non-CpG contexts [36].

Partnering of native proteins to accessory modules (MM_X) can be used to enhance editing or perturb the downstream steps of processing. This MM_X strategy has been most effectively used in genome base editing thus far, most notably in the use of AID/APOBEC enzymes for targeted editing of cytosine. The initial groundbreaking effort for genome editing employed APOBEC1 as the MM, along with dCas9, which plays a dual role in targeting and in unwinding of the target site so that the single-stranded DNA deaminase can catalyze a C>U transition. With this construct, in *cis* incorporation of uracil DNA glycosylase inhibitor (UGI) – a small phage-derived protein that potently inhibits UDG – prevents the downstream repair and greatly increases the efficiency of editing processes [37]. The strategy of coupling to UGI has been extended to constructs where AID homologs are used rather than APOBEC1 [38,39], and the bacteriophage μ protein Gam has similarly been employed as an MM_X to decrease indel frequencies [40]. This idea has also been applied to epigenome editing: fusing DNMT3L, which primarily acts as a catalytically inactive regulatory partner to DNMT3A *in vivo*, as an MM_X to DNMT3A enables multimerization and promotes processive methylation, which potentially helps to replicate the natural tendency for epigenetic modifications to cluster in large genic or intergenic regions [25]. It is feasible that the MM_X approach could be extended in the future by combining multiple MMs in *cis*; for example, locus-specific demethylation could be enhanced by fusing TDG to TET-containing constructs.

Modifying the enzymes (MM*)

Aside from harnessing native enzymes, existing structure-function knowledge has been exploited to generate variants of DNA-modifying enzymes with increased utility. Both rational and screening-based approaches have been utilized to assess and/or evolve activities, with the most significant advances coming in genome rather than epigenome editing thus far. For example, prior biochemical work had isolated determinants of sequence targeting for each AID/APOBEC family member [41], and targeted alteration to this region and others in the active site permitted increased precision in base editing [42]. Hyperactive DNA deaminases have also been identified with high-throughput methodologies [43,44], and some of these variants have been employed to increase gene editing efficiency or to introduce multiple diversifying edits to a localized region [45,46]. In the realm of epigenetic editing, the prokaryotic CpG methyltransferase M.SssI has recently been developed as a dCas9 fusion, and rational mutagenesis from computational modeling was used to decrease off-target effects [47].

Several recent biochemical insights into DNA-modifying enzymes have yet to be employed in concert with TMs but could likely find great utility. For example, it has proven very difficult to ascertain the independent roles of ox-mCs both *in vitro* and *in vivo*, as all three TET isozymes are known to make all three ox-mCs. Recently discovered hmC-dominant mutants of human TET2 [48] or *Naegleria* TET1 [35], or a TET homolog from the mushroom *Coprinopsis cinerea* that produces mostly fC [49], could be utilized in editing systems to more directly dissect the roles of individual ox-mCs. Similarly, while DNA deaminases have chiefly been used thus far for genome rather than epigenome editing, the discovery of AID variants with an increased specificity for 5mC, rather than unmodified cytosine, opens up the possibility of using AID-mediated deamination of 5mC to achieve DNA demethylation [50].

Spatiotemporal targeting control

To enhance spatial or temporal control over editing, small molecule modulators and additional TM or MM manipulations have been pursued, allowing for tunable activity of the modifying module (MM[†]). While small molecule-inducible promoters offer one simple solution to temporal control, more elegant solutions for spatiotemporal control have explored using specific protein-protein or protein-RNA interactions to bring together separate TM and MM constructs. One example involves tagging the MM, such as TET1 or a hyperactive AID, with the RNA-binding protein MS2, which is directed to dCas9 by incorporation of MS2 hairpin-binding sites into the gRNA [33,46], and similar strategies employ other specific RNA binding interactions such as PP7 [51]. Another example is the SunTag system, which relies on multiple antibody epitopes that can be recognized by a single-chain variable fragment (scFv). Fusing dCas9 with the epitopes and the MM to the scFv can result in multiple proteins being targeted to the designated locus, amplifying the activity [22,32]. Applying SunTag to DNMT3A, methylation was more efficiently spread over a region, similar to the effect seen by fusing DNMT3L as an MM_X [22]. In a different approach to achieving spatiotemporal control, fusing a TALE with cryptochrome-2 (Cry2)

interacting binding protein-1 (CIB1) and either TET1 or DNMT3A with Cry2 allowed for localization and epigenetic editing upon blue light illumination [27].

As splitting Cas9 to create a ligand-dependent activation variant has been an effective means to gain spatiotemporal control over TMs [52], an analogous protein-driven method for controlling MMs is to split the relevant enzyme so that it only reconstitutes at the desired site. A recent study using a split methyltransferase fused to dCas9 exhibited high specificity to target sites and suggested that the strategy limited off-target effects [53]. Although not yet attached to a targeting scaffold, TET2 has also been split and placed under the control of a small molecule modulator: Two inactive fragments were individually fused with either FKBP12 or FRB, and upon the addition of rapamycin, TET2 reassembled into a functional enzyme [54]. The approach of MM splitting has also not yet been applied to AID/APOBECs, but this would be anticipated to decrease off-target deamination outside the dCas9 DNA-binding motif.

Adenine modifications

Although genome and epigenome editing have so far focused on cytosine, several groups have used mass spectrometry and new sequencing technologies to highlight the potential versatility of other bases, especially adenine. For example, while N6-methyladenosine (m6A) is known to be quite abundant in prokaryotic DNA and eukaryotic RNA, recent studies have also identified m6A in the DNA of higher organisms [55–57], although some controversy exists about its presence in vertebrate genomes [58–61]. Preliminary studies have suggested a possible role for m6A in stem cell differentiation and transcriptional regulation, fueled by the identification and characterization of potential methyltransferase and demethylase enzymes in *C. elegans* [56]. Intriguingly, oxidation products of m6A (notably, hm6A and f6A) have been identified in RNA [62], so new DNA sequencing technologies may help discern whether these marks exist in DNA. In the future, to better understand the functional impact of adenine methylation, introducing or removing m6A at targeted loci may provide a means to overcome the low endogenous levels that, in part, obscure function. The parallel of adenine methylation and cytosine modification carries forward to genome editing as well. TadA is a tRNA adenine deaminase that typically converts adenosine to inosine in the single-stranded anticodon loop of tRNA. Evolution of TadA has recently achieved the remarkable feat of converting the enzyme to a DNA adenine deaminase, and fusion to dCas9 expanded the scope of editing to now include A:T to G:C changes [63].

Conclusions

The rapidity of recent progress highlights the tremendous potential of targeted genome and epigenome editing. Many challenges remain, such as delivering editing complexes to cells *ex vivo* or *in vivo* to access their therapeutic potential. Nonetheless, as we have discussed, these approaches promise to continue to reveal new biology, particularly if we harness current knowledge about the structure and function of the DNA-modifying enzymes in nature's toolbox.

Acknowledgments

We would like to acknowledge all the contributions that we are unable to cite in this brief review. We thank Junwei Shi for input. This work was supported by the Rita Allen Foundation (RMK), NIH (R01GM1185 and R21HG009545 to RMK; F30CA196097 to MYL) and NSF (EKS, JED).

References

1. Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*. 2017; 169:559.
2. Pulecio J, Verma N, Mejia-Ramirez E, Huangfu D, Raya A. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell*. 2017; 21:431–447. [PubMed: 28985525]
3. Nabel CS, Manning SA, Kohli RM. The Curious Chemical Biology of Cytosine: Deamination, Methylation, and Oxidation as Modulators of Genomic Potential. *ACS Chem Biol*. 2012; 7:20–30. [PubMed: 22004246]
4. Jurkowska RZ, Jurkowski TP, Jeltsch A. Structure and function of mammalian DNA methyltransferases. *Chembiochem*. 2011; 12:206–222. [PubMed: 21243710]
5. Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet*. 2017; 18:517–534. [PubMed: 28555658]
6. Drohat AC, Coey CT. Role of Base Excision “Repair” Enzymes in Erasing Epigenetic Marks from DNA. *Chem Rev*. 2016; 116:12711–12729. [PubMed: 27501078]
7. Siriwardena SU, Chen K, Bhagwat AS. Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases. *Chem Rev*. 2016; 116:12688–12710. [PubMed: 27585283]
8. Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol*. 2010; 11:607–620. [PubMed: 20683471]
9. Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol*. 2012; 8:751–758. [PubMed: 22772155]
10. Schutsky EK, Nabel CS, Davis AKF, DeNizio JE, Kohli RM. APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA. *Nucleic Acids Res*. 2017; 45:7655–7665. [PubMed: 28472485]
11. Nelson CE, Gersbach CA. Engineering Delivery Vehicles for Genome Editing. *Annu Rev Chem Biomol Eng*. 2016; 7:637–662. [PubMed: 27146557]
12. Hu JH, Davis KM, Liu DR. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. *Cell Chem Biol*. 2016; 23:57–73. [PubMed: 26933736]
13. Gersbach CA, Gaj T, Barbas CF 3rd. Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. *Acc Chem Res*. 2014; 47:2309–2318. [PubMed: 24877793]
14. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*. 2013; 14:49–55. [PubMed: 23169466]
15. Kubik G, Batke S, Summerer D. Programmable sensors of 5-hydroxymethylcytosine. *J Am Chem Soc*. 2015; 137:2–5. [PubMed: 25562518]
16. Zhang Y, Liu L, Guo S, Song J, Zhu C, Yue Z, Wei W, Yi C. Deciphering TAL effectors for 5-methylcytosine and 5-hydroxymethylcytosine recognition. *Nat Commun*. 2017; 8:901-017-00860-6. [PubMed: 29026078]
17. Rath P, Witte A, Summerer D. Engineering DNA Backbone Interactions Results in TALE Scaffolds with Enhanced 5-Methylcytosine Selectivity. *Sci Rep*. 2017; 7:15067-017-15361-1. This is the latest in a series of mutational studies that have refined the selectivity of TALE domains, in this case enhancing recognition of 5mC by reducing non-specific DNA backbone interactions. Thus, TALEs can increasingly be used to target specific cytosine modifications. [PubMed: 29118409]

18. Maurer S, Giess M, Koch O, Summerer D. Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine. *ACS Chem Biol.* 2016; 11:3294–3299. [PubMed: 27978710]
19. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol.* 2015; 13:722–736. [PubMed: 26411297]
20. Amabile A, Migliara A, Capasso P, Biffi M, Cittaro D, Naldini L, Lombardo A. Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell.* 2016; 167:219–232. e14. [PubMed: 27662090]
21. Bernstein DL, Le Lay JE, Ruano EG, Kaestner KH. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *J Clin Invest.* 2015; 125:1998–2006. [PubMed: 25866970]
22. Huang YH, Su J, Lei Y, Brunetti L, Gundry MC, Zhang X, Jeong M, Li W, Goodell MA. DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. *Genome Biol.* 2017; 18:176-017-1306-z. [PubMed: 28923089]
23. McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwards JR, Challen GA. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biol Open.* 2016; 5:866–874. [PubMed: 27170255]
24. Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, Rots MG, Ragozin S, Jurkowski TP, Jeltsch A. Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. *J Mol Biol.* 2013; 425:479–491. [PubMed: 23220192]
25. Stepper P, Kungulovski G, Jurkowska RZ, Chandra T, Krueger F, Reinhardt R, Reik W, Jeltsch A, Jurkowski TP. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res.* 2017; 45:1703–1713. By fusing DNMT3L to dCas9-targeted DNMT3A, the authors show that methylation can be enhanced at specific loci, including entire CpG islands, as the presence of DNMT3L and multimerization with DNMT3A promote processive methylation. [PubMed: 27899645]
26. Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, Klasic M, Zoldos V. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* 2016; 44:5615–5628. [PubMed: 26969735]
27. Lo CL, Choudhury SR, Irudayaraj J, Zhou FC. Epigenetic Editing of Ascl1 Gene in Neural Stem Cells by Optogenetics. *Sci Rep.* 2017; 7:42047. In a creative approach to introducing spatiotemporal control, the authors attach each part of their editing complex – the targeting module, TALE, and either TET1 or DNMT3A – to the protein pair CRY2 and CIB1. Upon blue-light illumination, the latter proteins dimerize and allow for the methylation or demethylation of the target CpGs. [PubMed: 28181538]
28. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE, et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol.* 2013; 31:1137–1142. [PubMed: 24108092]
29. Chen H, Kazemier HG, de Groote ML, Ruiters MH, Xu GL, Rots MG. Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Res.* 2014; 42:1563–1574. [PubMed: 24194590]
30. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. Editing DNA Methylation in the Mammalian Genome. *Cell.* 2016; 167:233–247. e17. This paper is one of the first to show the power and specificity of using CRISPR/dCas9 to target specific CpGs for methylation or demethylation by DNMT3A and TET1, respectively. [PubMed: 27662091]
31. Choudhury SR, Cui Y, Lubecka K, Stefanska B, Irudayaraj J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget.* 2016; 7:46545–46556. [PubMed: 27356740]
32. Morita S, Noguchi H, Horii T, Nakabayashi K, Kimura M, Okamura K, Sakai A, Nakashima H, Hata K, Nakashima K, Hatada I. Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol.* 2016; 34:1060–1065. In this

paper, the authors utilize an antibody-antigen binding system to bring TET1 to the DNA-targeted dCas9. This system allows multiple copies of scFv-TET1 to be recruited to SunTag-bound dCas9, increasing both the range of activity and methylation levels. [PubMed: 27571369]

33. Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, Ruan K, Wang F, Xu GL, Hu R. A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* 2016; 2:16009. [PubMed: 27462456]
34. Parrilla-Doblas JT, Ariza RR, Roldan-Arjona T. Targeted DNA demethylation in human cells by fusion of a plant 5-methylcytosine DNA glycosylase to a sequence-specific DNA binding domain. *Epigenetics.* 2017; 12:296–303. In this report, ROS1, a plant DNA glycosylase that directly excise 5mC, fused to the DNA-binding domain of yeast GAL4 (GBD) allows for targeted DNA demethylation and reactivation of specific silenced genes in human cells. [PubMed: 28277978]
35. Hashimoto H, Pais JE, Dai N, Correa IR Jr, Zhang X, Zheng Y, Cheng X. Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. *Nucleic Acids Res.* 2015; 43:10713–10721. [PubMed: 26323320]
36. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* 2012; 22:2497–2506. [PubMed: 22960375]
37. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature.* 2016; 533:420–424. Seminal work that introduced the concept of base editing by fusing dCas9 to an AID/APOBEC cytosine deaminase to catalyze site-specific targeted C>U editing. This work also introduces second and third generation base editors, which include the fusion of UGI and using nCas9 to address the challenge of off-target effects and aberrant downstream genomic changes. [PubMed: 27096365]
38. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science.* 2016; 353 Epub 2016 Aug 4. doi: 10.1126/science.aaf8729
39. Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, Qiu JL, Wang D, Gao C. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol.* 2017; 35:438–440. [PubMed: 28244994]
40. Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH, Liu DR. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv.* 2017; 3:eaao4774. [PubMed: 28875174]
41. Kohli RM, Maul RW, Guminski AF, McClure RL, Gajula KS, Saribasak H, McMahon MA, Siliciano RF, Gearhart PJ, Stivers JT. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J Biol Chem.* 2010; 285:40956–40964. [PubMed: 20929867]
42. Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol.* 2017; 35:371–376. [PubMed: 28191901]
43. Gajula KS, Huwe PJ, Mo CY, Crawford DJ, Stivers JT, Radhakrishnan R, Kohli RM. High-throughput mutagenesis reveals functional determinants for DNA targeting by activation-induced deaminase. *Nucleic Acids Res.* 2014; 42:9964–9975. [PubMed: 25064858]
44. Wang M, Yang Z, Rada C, Neuberger MS. AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity. *Nat Struct Mol Biol.* 2009; 16:769–776. [PubMed: 19543289]
45. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods.* 2016; 13:1029–1035. [PubMed: 27723754]
46. Hess GT, Fresard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods.* 2016; 13:1036–1042. [PubMed: 27798611]
47. Lei Y, Zhang X, Su J, Jeong M, Gundry MC, Huang YH, Zhou Y, Li W, Goodell MA. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nat Commun.* 2017; 8:16026. [PubMed: 28695892]

- 48•. Liu MY, Torabifard H, Crawford DJ, DeNizio JE, Cao XJ, Garcia BA, Cisneros GA, Kohli RM. Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. *Nat Chem Biol.* 2017; 13:181–187. As an important step toward dissecting the function of ox-mC bases, this work provides the first variants of human TET2 that largely stall oxidation at 5hmC. Identification of key active site interactions may also guide further manipulation of TET family enzyme activity. [PubMed: 27918559]
49. Zhang L, Chen W, Iyer LM, Hu J, Wang G, Fu Y, Yu M, Dai Q, Aravind L, He C. A TET homologue protein from *Coprinopsis cinerea* (CcTET) that biochemically converts 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. *J Am Chem Soc.* 2014; 136:4801–4804. [PubMed: 24655109]
50. Budzko L, Jackowiak P, Kamel K, Sarzynska J, Bujnicki JM, Figlerowicz M. Mutations in human AID differentially affect its ability to deaminate cytidine and 5-methylcytidine in ssDNA substrates in vitro. *Sci Rep.* 2017; 7:3873-017-03936-x. [PubMed: 28634398]
51. Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, Lim WA. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell.* 2015; 160:339–350. [PubMed: 25533786]
52. Oakes BL, Nadler DC, Flamholz A, Fellmann C, Staahl BT, Doudna JA, Savage DF. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat Biotechnol.* 2016; 34:646–651. [PubMed: 27136077]
- 53•. Xiong T, Meister GE, Workman RE, Kato NC, Spellberg MJ, Turker F, Timp W, Ostermeier M, Novina CD. Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. *Sci Rep.* 2017; 7:6732-017-06757-0. To improve specificity and limit off-target effects, these authors fuse the C-terminus of a DNA MTase to dCas9 and separately expressed the N-terminal portion. [PubMed: 28751638]
- 54•. Lee M, Li J, Liang Y, Ma G, Zhang J, He L, Liu Y, Li Q, Li M, Sun D, et al. Engineered Split-TET2 Enzyme for Inducible Epigenetic Remodeling. *J Am Chem Soc.* 2017; 139:4659–4662. In this report, fusing two inactive fragments of a split-TET2 with FKBP12 or FRB permitted rapamycin-inducible reassembly into a functional enzyme. [PubMed: 28294608]
55. Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, Hao Z, Liu J, Lu X, Dore LC, et al. N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell.* 2015; 161:879–892. [PubMed: 25936837]
56. Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, Hsu CH, Aravind L, He C, Shi Y. DNA Methylation on N6-Adenine in *C. elegans*. *Cell.* 2015; 161:868–878. [PubMed: 25936839]
- 57•. Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, Yin R, Zhang D, Zhang P, Liu J, et al. N6-methyladenine DNA modification in *Drosophila*. *Cell.* 2015; 161:893–906. This work provides the first evidence of significant m6A modification in *Drosophila* genomes using ultra-high-performance LC-MS/MS techniques. This study also identified a TET homologue that displayed m6A demethylase activity, and showed that it has a role in embryonic development. [PubMed: 25936838]
58. Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, Liu Y, Byrum SD, Mackintosh SG, Zhong M, et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature.* 2016; 532:329–333. [PubMed: 27027282]
59. Koziol MJ, Bradshaw CR, Allen GE, Costa ASH, Frezza C, Gurdon JB. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat Struct Mol Biol.* 2016; 23:24–30. [PubMed: 26689968]
60. Schiffrers S, Ebert C, Rahimoff R, Kosmatchev O, Steinbacher J, Böhne AV, Spada F, Michalakis S, Nickelsen J, Müller M, Carell T. Quantitative LC-MS Provides No Evidence for m6 dA or m4 dC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angew Chem Int Ed Engl.* 2017; 56:11268–11271. [PubMed: 28371147]
61. Liu J, Zhu Y, Luo GZ, Wang X, Yue Y, Wang X, Zong X, Chen K, Yin H, Fu Y, et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat Commun.* 2016; 7:13052. [PubMed: 27713410]

62. Fu Y, Jia G, Pang X, Wang RN, Wang X, Li CJ, Smemo S, Dai Q, Bailey KA, Nobrega MA, et al. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat Commun.* 2013; 4:1798. [PubMed: 23653210]
- 63••. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. Programmable base editing of A:T to G:C in genomic DNA without DNA cleavage. *Nature*. 2017 advance online publication. Using directed evolution and protein engineering, the authors evolved a tRNA adenosine deaminase into a DNA-editing enzyme, which can convert A:T to G:C base pairs in living cells. This marks a critical advance in expanding the scope of targeted, direct genome editing and carries potential therapeutic applications.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- Site-specific genome and epigenome editing are rapidly advancing
- Knowledge of DNA-modifying enzymes can be exploited to improve activity
- Editing complexes can be used to directly probe the biology of DNA modifications

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Nature's 'toolbox' for cytosine modifications

The toolbox for targeted genome modifications includes many enzymes that act on cytosine bases. Cytosine can be methylated by a DNA methyltransferase (DNA MTase, green) to generate 5mC, which, in turn, can be oxidized by the TET enzymes (blue) to 5hmC, 5fC, and 5caC. TDG is a glycosylase that then removes 5fC and 5caC and generates an abasic site (purple) for base excision repair. Cytosine can also be deaminated by the AID/APOBEC enzymes (red) to form uracil, which is subject to excision by UDG, a repair pathway that can be inhibited by UGI. The Repressor of Silencing 1 (ROS1) enzyme found in plants can be used to excise 5mC and generate an abasic site. Not shown are some possible crosstalk between modifying enzymes, such as the deamination of 5mC by AID/APOBEC enzymes.

Figure 2. The formula for editing complexes and ongoing developments

Editing complexes can be described as consisting of a Targeting Module (TM) and Modifying Module (MM). As highlighted in the text, characterizing the complexes in this manner allows for the various forms to be classified based on other distinguishing features, including the addition of accessory modules (MM_x), variants of the native enzymes with altered activities (MM*), or tunable variants with altered spatiotemporal control (MM[†]). Established examples of these complexes are shown, as well as anticipated advances that are based on knowledge of the structure and function of the DNA-modifying enzymes. Lines connecting the modules represent fusions or linkers, and a plus sign indicates that the two relative entities were co-expressed separately. The dashed line indicates that both fused and unfused/co-expressed versions have been made.

Figure 3. Nature's 'toolbox' for adenine modifications

Analogous to cytosine modification, adenine can be methylated on the exocyclic amine by a methyltransferase (MTase, green), which can then be oxidized (blue) by TET-like oxygenase enzymes (e.g. FTO) to promote demethylation. Oxidized m6A bases can be stable in RNA, raising the possibility that such bases might exist or be introduced in DNA (dashed line). Adenine can also be deaminated to form inosine (red) using a variant of the tRNA editing enzymes, TadaA, evolved to act on DNA. Inosine is the subject of DNA repair in bacteria (purple) and mammals. Other potential pathways for repair enzymes to interact with m6A or other adenine variants may also exist in cells (dashed lines).